CHANGE ANALYSIS AND FISHER-SCORE CHANGE PROCESSES

Emanuel Parzen

Department of Statistics

Texas A&M University

Technical Report No. #171

May 1992

Texas A&M Research Foundation

Project No. 6547

'Functional Statistical Data Analysis and Modeling'

Sponsored by the U. S. Army Research Office

Professor Emanuel Parzen, Principal Investigator

REPORT DOCUMENTATION PAGE.

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1992	

4. TITLE AND SUBTITLE
CHANGE ANALYSIS AND FISHER-SCORE PROCESSES

5. FUNDING NUMBERS
DAAL03-90-G-0069

6. AUTHOR(S)
EMANUEL PARZEN

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Texas A&M University
Department of Statistics
College Station, TX 77843-3143

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 27574.7-MA

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
This paper aims to synthesize classical statistical methods and changepoint hypothesis testing and to contribute to solutions of the historical basic applied problem of statistics: distinguish change (of the model) from fluctuation (within the model), the variability expected under homogeneity. Contents are: 0. Goals, 1. Comparison change analysis as probability study of (X,Y); 2. Asymptotic distributions of sample change processes; 3. One way analysis of variance (AOV); 4. Change analysis approach to AOV; 5. Components of change analysis; 6. Four phases of change analysis; 7. Nonparametric statistics fro multisample analysis; 8. Fisher-Score change processes.

14. SUBJECT TERMS
change density, change process, comparison density, comparison distributions, Fisher-Score change densities, empirical change processes theory, components of change processes, phases of change analysis.

15. NUMBER OF PAGES
18

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

May 21, 1992

CHANGE ANALYSIS AND FISHER-SCORE CHANGE PROCESSES

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, Texas 77843-3143 USA

This paper is written for discussion at the AMS-IMS-SIAM Summer Research Workshop on "Changepoint Analysis", at Mt. Holyoke College, July 11-16, 1992 and at the Carleton University Workshop on "Changepoint Analysis," August 31-September 5, 1992. Contents are:

0. Goals
1. Comparison change analysis as probability study of (X, Y);
2. Asymptotic distributions of sample change processes;
3. One way analysis of variance (AOV);
4. Change analysis approach to AOV;
5. Components of change analysis;
6. Four phases of change analysis;
7. Nonparametric statistics for multisample analysis;
8. Fisher-Score change processes.

Change last year,
Change the year before!
Expect Change this year,
Unlike any change of yore?
To detect change, without fear,
CUSUM process your score . . .
Unity makes practice of statistics clear.
Who could ask for anything more?

0. GOALS

Ultimate goals of our research program include: unify parametric and nonparametric inference for continuous and discrete data; synthesize classical statistical methods and changepoint hypothesis testing; demonstrate that mathematical statistical and data analytic approaches are both needed for statistical inference; stimulate exoteric methods (applicable by applied researchers) rather than esoteric methods (known only to a small group of mathematical statisticians); combine mathematical statistical and data analytic views to develop methods of statistical analysis which are based on assumptions (known model) which are tested in ways that provide insight how to model deviations of the data from the assumed model (and thus often identify a "true" model as an "iterated" model which models "residuals"); contribute to solutions of the historical basic applied problem of statistics: distinguish change (of the model) from fluctuation (within the model), the variability expected under homogeneity.

This paper is not a finished or rigorous presentation of results; it is a stimulus for discussion about open research problems in change analysis. One need may be to determine how to develop a classification scheme to catalog the past and future extensive literature about statistical methods to model change.

Research supported by the U. S. Army Research Office

1. COMPARISON CHANGE ANALYSIS AS PROBABILITY STUDY OF (X,Y)

This section outlines the notation and concepts that we introduce (Parzen (1992)) in our probability theory of the relations between two random variables X and Y. They motivate the statistics that we propose to describe the changes over time of a series of observations $Y(t)$, $t = 1, 2, \ldots$. To apply the probability theory of (X,Y) to data, let X represent t, the index of observation.

The distribution function, quantile function, probability mass function, and probability density functions of Y are respectively denoted $F_Y(y)$, $Q_Y(u)$, $p_Y(y)$, $f_Y(y)$. *We assume that Y is either discrete or continuous, X is either discrete or continuous.*

To develop a theory that applies to both discrete and continuous variables we define τ, $0 < \tau < 1$, to be an X-exact value if

$$F_X(Q_X(\tau)) = \tau.$$

If $F_X(\cdot)$ is continuous, all τ are X-exact. If $F_X(\cdot)$ is discrete, τ is X-exact if there exists value x at which F_X jumps and $x = Q_X(\tau)$ (therefore $F_X(x) = \tau$).

Let U denote a random variable which is Uniform[0,1]. If Y is continuous, the probability integral transform $F_Y(Y)$ is identically distributed as U. If Y is discrete we transform Y to

$$F_Y^{mid}(Y) = F_Y(Y) - .5p_Y(Y).$$

If u is Y-exact,

$$\text{Prob}[F_Y^{mid}(Y) \leq u] = u = \text{Prob}[Y \leq Q_Y(u)].$$

A function $J(u), 0 < u < 1$, is called a score function (to be more precise, Y-score function); it is called normalized if

$$\int_0^1 J(u) = 0, \int_0^1 J^2(u)du = 1.$$

Score change density and score change process: Define for $0 < \tau < 1$

$$c(\tau, J) = E[J(F_Y^{mid}(Y))|X = Q_X(\tau)]] - E[J(U)],$$
$$C(\tau, J) = \int_0^\tau c(t, J)dt.$$

For a sequence $Y(t), t = 1, \ldots, n$, analogous concepts are, defining

$$Y^- = (1/n)\sum_{t=1}^n Y(t), \quad f^- = (1/n)\sum_{t=1}^n f(Y(t)),$$

the sample change density

$$c^-(\tau) = Y(t) - Y^-, (j-1)/n < \tau < j/n, j = 1, \ldots, n.$$

and the sample change process

$$C^-(\tau) = \int_0^\tau c^-(t)dt, 0 < \tau < 1$$

which is our version of CUSUMS.

Patterns in these change processes will be examined by computing linear functionals for suitable change-score functions $K(\tau), 0 < \tau < 1$. Define

$$[c(.,J),K] = [J,K] = \int_0^1 c(\tau,J)K(\tau)d\tau$$

We call $[J,K]$ a *double score* component. It measures how $c(\tau,J)$ behaves as a function of τ (for J fixed).

Change Theorem C: $C(\tau,J)$ linearly interpolates its values at X-exact values of τ, where it satisfies

$$C(\tau,J) = \tau(E[J(F_Y^{mid}(Y))|X \leq Q_X(\tau)]] - E[J(U)]).$$

The proof of Change Theorem C requires the methodology (Parzen (1979), (1991), (1992), (1993)) of comparison density functions $d(u;F,G)$ and comparison distributions $D(u;F,G)$; they compare two distributions F and G which are either discrete or continuous. $D(u)$ is defined as the integral of $d(u)$, $d(u) = D'(u)$. When $d(u)$ is piecewise constant, $D(u)$ is piecewise linear. When F and G are both continuous we define $D(u;F,G) = G(F^{-1}(u))$.

Change Dependence Densities and Distributions: define, for $0 < \tau, u < 1$

$$d(\tau,u) = d(u;F_Y,F_{Y|X=Q_X(\tau)}),$$
$$d([0,\tau],u) = d(u;F_Y,F_{Y|X\leq Q_X(\tau)}).$$

$$D(\tau,u) = D(u;F_Y,F_{Y|X=Q_X(\tau)}),$$
$$D([0,\tau],u) = D(u;F_Y,F_{Y|X\leq Q_X(\tau)}).$$

Best Change Theorem D: For τ X-exact and u Y-exact

$$\tau d([0,\tau],u) = \int_0^\tau d(t,u)dt$$

We call this theorem best because it explains why estimators of $\tau d([0,\tau],u)$ for fixed τ have asymptotic variances similar to that of probabilities rather than densities, and it yields proofs of all change theorems stated. The proof of Change Theorem D is outlined in Parzen (1992).

Change Theorem E: $c(\tau,J) = \int_0^1 J(u)(d(\tau,u) - 1)du$

$$C(\tau,J) = \tau \int_0^1 J(u)(d([0,\tau],u) - 1)du$$
$$[J,K] = \int_0^1 \int_0^1 K(\tau)J(u)(d(\tau,u) - 1)dud\tau$$

Important score functions are *indicator* score functions $J(.;u)$: $J(u';u) = 1$ or 0 as $u' \leq u$ or $u' > u$. Assume u is Y-exact. Denote by $c(\tau,u)$ and $C(\tau,u)$ the change density and change process of $J(.;u)$:

$$c(\tau,u) = Prob[F_Y^{mid}(Y) \leq u|X = Q_X(\tau)]] - u,$$
$$C(\tau,u) = \tau(Prob[F_Y^{mid}(Y) \leq u|X \leq Q_X(\tau)]] - u).$$

Change Theorem F: At X-exact τ and Y-exact u

$$C(\tau, u) = \tau(D([0, \tau], u) - u)$$
$$C(\tau, u) = D(\tau, u) - \tau u$$

Another important score function is $J(u) = Q_Y(u)$. Its change density and change process correspond to *conditional means* of Y:

$$c(\tau, Q_Y) = E[Y|X = Q_X(\tau)] - E[Y],$$
$$C(\tau, Q_Y) = \tau(E[Y|X \leq Q_X(\tau)] - E[Y])$$

Measures of dependence can be defined in terms of

$$\int_0^1 |c(\tau, Q_Y)|^2 d\tau = \text{VAR}(E[Y|X]),$$

$$\int_0^1 C(\tau, Q_Y) d\tau = \int_0^1 -(s - .5)c(s, Q_Y) ds$$

When X and Y are jointly normal with correlation coefficient ρ,

$$E[Y - E[Y]|X] = (\sigma[Y]/\sigma[X])\rho(X - E[X])$$

Therefore the change density of Y given X, when (X, Y) is bivariate normal, is

$$c(\tau, Q_Y) = \sigma[Y]\rho\Phi^{-1}(\tau).$$

Its integral is $C(\tau, Q_Y)$ whose graph has the typical shape of a change process which is able to detect whether there is a change in Y as a function of X.

To test the independence of X and Y one examines change processes $c(\tau, g(Q_Y))$ for several transformations g, which correspond to conditional means of non-linear functions of Y. *The problem in practice is how to choose informative non-linear functions.*

If one assumes a parametric model $f_\theta(y)$ for the true density $f_Y(y)$, where θ is a vector parameter with components θ_j, one choose non-linear functions of Y equal to *Fisher-score* functions, defined by

$$S_j(y, \theta) = (\partial/\partial\theta) log f_\theta(y).$$

Fisher-score change densities are defined to be $c(\tau, S_j(Q_Y, \theta))$.

They are called parametric change densities in contrast with $c(\tau, J)$ which are non-parametric change densities.

2. ASYMPTOTIC DISTRIBUTIONS OF SAMPLE CHANGE PROCESSES

Empirical process theory studies limit theorems for

$$C^{\sim}(f) = \frac{1}{n}\sum_{t=1}^{n}(f(Y(t)) - E[f(Y(t))])$$

uniform in f belonging to a specified family of functions $f(y)$.

Empirical change processes theory studies limit theorems for

$$C^{\sim}(\tau, f) = \frac{1}{n}\sum_{t=1}^{[n\tau]}(f(Y(t)) - f^{-})$$

where $f^{\sim}$ is the sample mean of $f(Y(t))$.

Sample versions of change processes $C(\tau, J)$ and $C(\tau, u)$ computed from a sample (of size n) are denoted $C^{\sim}(\tau, J)$ and $C^{\sim}(\tau, u)$; using theorems in the literature (especially Csörgő and Horvath (1987)) one can show that they have large sample asymptotic distributions under the hypothesis of no change (where $B(\tau)$ is a Brownian Bridge and $B(\tau, u)$ is a Brownian sheet)

$n^{.5}C^{\sim}(\tau, Q^{\sim}_Y)$, $0 < \tau < 1$ converges to $B(\tau)$, $0 < \tau < 1$.

$n^{.5}C^{\sim}(\tau, J), 0 < \tau < 1$ converges to $B(\tau), 0 < \tau < 1$, assuming J normalized,

$n^{.5}C^{\sim}(\tau, u), 0 < \tau, u < 1$ converges to $B(\tau, u), 0 < \tau, u < 1$;

for fixed τ, $(n/\tau(1-\tau))^{.5}C^{\sim}(\tau, u), 0 < u < 1$ converges to $B(u), 0 < u < 1$.

Parzen and Horvath (research in progress) establish similar asymptotic theorems for Fisher-score change processes

$$n^{.5}C^{\sim}(\tau, S_j(Q_Y, \hat{\theta}))$$

where $\hat{\theta}$ is a maximum likelihood estimator.

For comparison distributions, the asymptotic distributions under the null hypothesis of no change are

$$n^{.5}(D^{\sim}(\tau, u) - \tau u) \text{ converges to } B(\tau, u)$$

$$n^{.5}\tau(D^{\sim}([0, \tau], u) - u) \text{ converges to } B(\tau, u)$$

The Pyke-Shorack two sample process can be expressed: for fixed τ, as n tends to ∞,

$$(n\tau/(1-\tau))^{.5}(D^{\sim}([0, \tau], u) - u) \text{ converges to } B(u).$$

The sample distribution function $F^{\sim}$ of a sample (of size n_1) from true distribution F can be studied as the limit of two samples as $\tau \to 0$, first sample size $n_1 = n\tau \to \infty$; empirical processes can be expressed

$$n_1^{.5}(D^{\sim}(u; F, F^{\sim}) - u) \text{ converges to } B(u).$$

The foregoing are asymptotic distributions of sample change processes under the *null hypothesis of no change*. Of great interest are their asymptotic distributions under *alternative hypotheses of change*.

The sample comparison function $D(u; G, F^{\sim})$ of the sample distribution function $F^{\sim}$ with a model G, when the sample of size n_1 has true distribution function F, has asymptotic distribution

$$n_1^{.5}\left(D\left(u; G, F^{\sim}\right) - D\left(u; G, F\right)\right) \to B_F\left(D\left(u; G, F\right)\right)$$

where

$$B_F(u) = n_1^{.5}(D(u; F, F^{\sim}) - u)$$

is called the empirical process of the sample and is approximately a Brownian Bridge.

Under suitable conditions

$$n_1^{.5}(D^{-1}(u; G, F^{\sim}) - D^{-1}(u; G, F))$$
$$\to \left(\frac{d}{du}D^{-1}(u; G, F)\right)(-1)B_F(u)$$

The comparison of the sample up to time τ (of size $n\tau$) with the whole sample (of size n), under the changepoint assumption that the sample up to time τ and the sample after time τ have respective true distributions $F([0, \tau], y)$ and $F([\tau, 1], y)$ and pooled sample

has distribution $F_Y(y) = F([0,1], y)$, has asymptotic distribution for fixed τ suggested by Pyke-Shorack theory for two samples: as processes on $0 < u < 1$,

$$\begin{aligned} & n^{.5}\tau\left(D\tilde{}([0,\tau],u) - D([0,\tau],u)\right) \\ & \rightarrow \tau d([0,\tau],u)B_{[\tau,1]}((1-\tau)D([\tau,1],u)) \\ & -(1-\tau d([0,\tau],u)B_{[0,\tau]}(\tau D([0,\tau],u)) \end{aligned}$$

where $B_{[0,\tau]}(u)$ and $B_{[\tau,1]}(u)$ are the empirical processes of the samples before and after τ respectively. Note that $B(\tau D)$ symbolizes $\tau^{.5}B(D)$, and

$$\tau D([0,\tau],u) + (1-\tau)D([\tau,1],u) = u.$$

From Ruymgaart (1974) we obtain results when (X,Y) has a continuous bivariate distribution:

$$\begin{aligned} & n^{.5}\Big(\int_0^1\int_0^1 K(\tau)J(u)dD\tilde{}(\tau,u) - \int_0^1\int_0^1 K(\tau)J(u)dD(\tau,u)\Big) \\ & \rightarrow \text{ Normal } (0)\int_0^1\int_0^1 |V(\tau,u)|^2 dD(\tau,u)) \end{aligned}$$

defining

$$\begin{aligned} V(\tau,u) = K(\tau)J(u) & - \int_0^1\int_0^1 K(t)J(s)dD(t,s) \\ & + \int_0^1\int_0^1 K(t)[e(s-u)-s]J'(s)dD(t,s) \\ & + \int_0^1\int_0^1 K'(t)[e(t-\tau)-t]J(s)dD(t,s) \end{aligned}$$

where $e(x) = 1$ or 0 as $x \geq 0$ or $x < 0$. Under the null hypothesis that X and Y are independent, $D(\tau,u) = \tau u$,

$$V(\tau,u) = (K(\tau) - \int_0^1 K(t)dt)(J(u) - \int_0^1 J(s)ds),$$

and

$$n^{.5}(D\tilde{}(\tau,u) - D(\tau,u)) \rightarrow B(\tau,u).$$

Note that (Weiss (1964))

$$\text{Cov}\left[n^{.5}f_X Q_X(\tau)\left\{Q\tilde{}_X(\tau) - Q_X(\tau)\right\},\quad n^{.5}f_Y Q_Y(u)\left\{Q\tilde{}_Y(u) - Q_Y(u)\right\}\right]$$

is aymptotically $D(\tau,u) - \tau u$.

3. ONE WAY ANALYSIS OF VARIANCE (AOV)

Change analysis provides new graphical data analysis interpretations of classical statistical methods. The one way analysis of variance (AOV) tests the equality of distributiions of variables (or populations) $Y_1, \ldots, Y_c$ under the assumption that they are independent and their distributions satisfy

$$Y_j \text{ is Normal}(\mu_j, \sigma^2), j = 1, \ldots, c.$$

Note that if one has observed values $Y(t), t = 1, \ldots, n$, of a variable Y, the variables Y_j could represent the values of $Y(t)$ for the j-th time segment $T_{j-1} < t \leq T_j$, where $0 = T_0 < T_1 < \ldots < T_c = n$ are specified by the statistician.

The parameters to be estimated are $\mu_1, \ldots, \mu_c, \sigma$. The basic hypothesis to be tested is the hypothesis of homogeneity

$$H_0 : \mu_1 = \ldots = \mu_c = \mu.$$

For $j = 1, \ldots, c$, one observes n_j values of Y_j denoted $Y_{j1}, \ldots, Y_{jn_j}$ with sample mean

$$\bar{Y}_{j\cdot} = (1/n_j) \sum_{i=1}^{n_j} Y_{ji}$$

and sample variance

$$S_j^2 = (1/n_j) \sum_{i=1}^{n_j} (Y_{ji} - \bar{Y}_{j\cdot})^2.$$

The *pooled sample* of all the data has size $n = n_1 + \ldots + n_c$. The proportion of the pooled sample from the j-th sample is

$$p_j = n_j/n;$$

the cumulative proportions are denoted

$$\tau_j = p_1 + \ldots + p_j.$$

We introduce a variable X to represent the population $j = 1, \ldots, c$ from which an observation Y_{ji} is made. An observation is (X, Y). The sample probability and distribution of X is

$$p_{\bar{X}}(j) = p_j, F_{\bar{X}}(j) = \tau_j.$$

The variable X is not a random variable, but we condition Y by X using sample (empirical) probabilities rather than population (ensemble) probabilities. We find it an aid to understanding to use an alternate notation for $\bar{Y}_{j\cdot}$ and S_j^2 as the sample conditional mean and variance of Y given $X = j$:

$$\bar{E}[Y|X = j] = \bar{Y}_{j\cdot}.$$
$$\bar{\mathrm{VAR}}[Y|X = j] = S_j^2.$$

The pooled sample has sample mean $\bar{Y}_{\cdot\cdot}$ and sample variance S_Y^2 which can be interpreted as unconditional mean and variance of Y:

$$\bar{E}[Y] = \bar{Y}_{\cdot\cdot} = \sum_{j=1}^{c} p_j \bar{Y}_{j\cdot}.$$

$$\bar{VAR}[Y] = S_Y^2 = (1/n) \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ji} - \bar{Y}_{\cdot\cdot})^2$$

The theory of conditional expectation has important formulas

$$\mathrm{VAR}[Y] = \mathrm{VAR}[E[Y|X]] + E[\mathrm{VAR}[Y|X]].$$
$$\mathrm{VAR}^{\sim}[Y] = \mathrm{VAR}^{\sim}[E^{\sim}[Y|X]] + E^{\sim}[\mathrm{VAR}^{\sim}[Y|X]].$$

Analysis of variance tests H_0 by comparing various estimators of variance. The mean conditional variance (denoted S^2_{var}) and the variance of the conditional mean (denoted S^2_{mean}) are defined by

$$S^2_{\mathrm{var}} = E^{\sim}[\mathrm{VAR}^{\sim}[Y|X]] = \sum_{j=1}^{c} p_j S_j^2,$$

$$S^2_{\mathrm{mean}} = \mathrm{VAR}^{\sim}[E^{\sim}[Y|X]] = \sum_{j=1}^{c} p_j (Y^{\sim}_{j.} - Y^{\sim}_{..})^2$$

The traditional F test statistic (denoted FT) for testing H_0 can be represented

$$FT = ((n-c)/(c-1))F,$$
$$F = S^2_{\mathrm{mean}}/S^2_{\mathrm{var}}$$

An estimator of σ^2 in the AOV model is

$$S^2 = (n/(n-c))S^2_{\mathrm{var}};$$

it is unbiased since $E[S^2] = \sigma^2$. The numerator of the F statistic can be shown to have mean

$$E[S^2_{\mathrm{mean}}] = (c-1)\sigma^2 + \sum_{j=1}^{c} p_j(\mu_j - \mu)^2,$$

defining

$$\mu = \sum_{j=1}^{c} p_j \mu_j.$$

The numerator and denominator of F can be shown to be independent random variables; therefore

$$E[FT] = 1 + (c-1)^{-1} \sum_{j=1}^{c} p_j((\mu_j - \mu)/\sigma)^2$$

This formula for the mean of FT is used to justify why we should reject the hypothesis H_0 of equal means when FT is too large; $FT > 2$ is a reasonable general criterion for rejecting H_0. Akaike (1985) describes the emergence of the magic number 2. The critical value of FT is exactly determined from the fact it obeys an F distribution with $(c-1, n-c)$ degrees of freedom.

Data analysis by analysis of variance is usually presented as an AOV table.

4. CHANGE ANALYSIS APPROACH TO AOV

The change analysis approach to AOV provides graphical analysis of the standardized data

$$Y^* = (Y - E^{\sim}[Y])/S_Y.$$

by forming processes on the unit interval $0 < \tau < 1$ defined as follows;
change density: $c\tilde{}(\tau) = Y^{*\tilde{}}_j = (Y\tilde{}_{j.} - Y\tilde{}_{..})/S_Y, \tau_{j-1} < \tau < \tau_j$;
change process: $C\tilde{}(\tau) = \int_0^\tau c\tilde{}(s)ds$;
change test process: $CT\tilde{}(\tau) = C\tilde{}(\tau)/(\tau(1-\tau))^{.5}$;
change test density: $cT\tilde{}(\tau) = c\tilde{}(\tau)(p_j/(1-p_j)^{.5}, \tau_{j-1} < \tau \leq \tau_j$.
The change process is linear between its values at $\tau = \tau_j$:

$$C\tilde{}(\tau_j) = \sum_{i=1}^{j} p_i(Y\tilde{}_{i.} - Y\tilde{}_{..})/S_Y = \tau_j E\tilde{}[Y^*|X \leq j].$$

The process $n^{.5}C\tilde{}(\tau)$, $0 < \tau < 1$, can be shown to have an asymptotic distribution under H_0 at the "exact" values $0 = \tau_0 < \tau_1 < \ldots < \tau_c = 1$ which is the same as the distribution of a Brownian Bridge stochastic process $B(\tau)$, $0 < \tau < 1$, a zero mean Gaussian process with covariance kernel

$$E[B(\tau_1)B)\tilde{}_2] = min(\tau_1, \tau_2) - \tau_1\tau_2.$$

We call $C\tilde{}(\tau)$ a *dynamic statistic* since the significance of its graph can be determined by thinking of it as a sample path of a Brownian Bridge. We also relate its graph to various deterministic shapes it could have under various assumptions about the values of the means μ_j.

Graphical data analysis of $C\tilde{}(.)$ can often indicate whether to accept or reject H_0. To obtain "p values" for the level at which H_0 is rejected or accepted we need to form functionals of the process.

Theorem: The important functional

$$R^2 = \int_0^1 |c\tilde{}(\tau)|^2 d\tau$$

can be related to the traditional F test statistic FT by

$$FT = ((n-c)/(c-1))F, F = R^2/(1-R^2).$$

Proof: Verify that

$$R^2 = S^2_{\text{mean}}/S_Y,$$
$$S_Y = S^2_{\text{mean}} + S^2_{\text{var}},$$
$$RV = 1 - R^2 = S^2_{\text{var}}/S^2_Y.$$

The distribution of R^2 under H_0 is analogous to sample correlation; therefore we call R^2 a correlation statistic to distinguish it from an F statistic of the form $F = R^2/(1-R^2)$.

F tests (and R^2 tests) are "portmanteau" statistics which should be represented in terms of diagnostic statistics which can help indicate which part of the data is the cause of rejection of the null hypothesis. For this purpose we introduce "two sample statistics" for the no-change hypotheses

$H_{j<}$: The pooled sample of variables $Y_1, \ldots Y_j$ has same distribution as the pooled sample of variables $Y_{j+1}, \ldots, Y_c$,

$H_{j=}$: The variable Y_j has the same distribution as the pooled sample which does not include Y_j.

Denote by $TT_{j<}$ the two sample t-test statistic for $H_{j<}$ and denote by $TT_{j=}$ the two sample t-test statistic for $H_{j=}$. One can show

$$TT_{j=} = ((n-c)p_j/(1-p_j))^{.5}(Y^-_{j.} - Y^-..)/S_Y$$
$$TT_{j<} = ((n-c)\tau_j/(1-\tau_j))^{.5}(E^-[Y|X \leq j] - Y^-..)/S_Y.$$

Therefore

$$TT_{j<} = ((n-c)/(1-R^2))^{.5}CT^-(\tau_j),$$
$$TT_{j=} = ((n-c)/(1-R^2))^{.5}cT^-(\tau_j)$$

The portmanteau F test statistic can be expressed

$$FT = (c-1)^{-1}\sum_{j=1}^{c}(1-p_j)|TT_{j=}|^2.$$

5. COMPONENTS OF CHANGE PROCESSES

We call the two-sample t statistics TT "abrupt change" statistics since they test hypotheses of an abrupt change. We would like statistics that test for smooth change (such as linear or quadratic). Natural test statistics are linear functionals in the change density process, called components $T(K)$ or $[c^-, K]$ with score function $K(\tau)$, defined by

$$T(K) = [c^-, K] = \int_0^1 K(\tau)c^-(\tau)d\tau = \sum_{j=1}^{c} Y^{*-}{}_j \int_{\tau_{j-1}}^{\tau_j} K(\tau)d\tau$$

The identity score function $K(\tau) = \tau$ yields the Wilcoxon type statistic

$$[c^-, \tau] = \sum_{j=1}^{c} Y^{*-}{}_j \; p_j \; .5(\tau_{j-1} + \tau_j)$$

A general approximation for a component is

$$[c^-, K] == \sum_{j=1}^{c} Y^{*-}{}_j p_j K(\tau_j^{mid})$$

defining

$$\tau_j^{mid} = .5(\tau_{j-1} + \tau_j) = \tau_j - .5p_j$$

To express the statistics $TT_{j=}$ and $TT_{j<}$ as components we first define score functions

$$K_{j=}(\tau) = 1/p_j \text{ for } \tau_{j-1} < \tau < \tau_j, = 0, \text{ otherwise;}$$
$$K_{j<}(\tau) = 1 \text{ for } 0 < \tau < \tau_j, = 0, \text{ otherwise.}$$

It can be shown that under H_0 a component $T(K)$ is asymptotically normal with mean 0 and variance

$$\text{Norm}(K)^2 = \int_0^1 |K(\tau) - \int_0^1 K(s)ds|^e d\tau$$

The identity score function $K(\tau) = \tau$ has norm squared 1/12. Therefore the asymptotically Normal (0,1) version of the Wilcoxon type test statistic is the component $T(12^{.5}\tau)$.

The score functions corresponding to the TT statistics have norms square

$$\text{Norm}(K_{j<})^2 = \tau(1-\tau),$$
$$\text{Norm}(K_{j=})^2 = (1-p_j)/p_j$$

Consequently one can represent the two sample t statistics as components

$$TT_{j<} = T(K_{j<}/\text{Norm}(K_{j<}))$$
$$TT_{j=} = T(K_{j=}/\text{Norm}(K_{j=}))$$

6. FOUR PHASES OF CHANGE ANALYSIS

A sample change process $C\tilde{}(\tau), 0 < \tau < 1$, is a dynamic statistic (sample path of a stochastic process) which often can be shown to satisfy under the null hypothesis of "no change" the null hypothesis $H_0 : C\tilde{}(.)$ is a Brownian Bridge (or a related hypothesis). The statistical analysis of $C\tilde{}(.)$ has four phases:

Phase 1: *Graphical analysis*; is the plot of $C\tilde{}(\tau)$, $0 < \tau < 1$, oscillatory, a deterministic parabola, other patterns.

Phase 2: *Non-linear functionals*. One tests H_0 by computing the values of test statistics (whose asymptotic distributions under H_0 can be deduced from the theory of empirical processes)

$$\int_0^1 |C\tilde{}(\tau)|^2 d\tau,$$
$$\int_0^1 (|C\tilde{}(\tau)|^2/\tau(1-\tau))d\tau,$$
$$\max_{0<\tau<1} |C\tilde{}(\tau)|,$$
$$\max_{\tau=j/n} |C\tilde{}(\tau)|/\tau(1-\tau).$$

Phase 3: *Linear functionals*. For various score functions $K(\tau)$, called *change score* functions, one computes the linear functional (or component)

$$C\tilde{}(K) = \int_0^1 K(\tau) dC\tilde{}(\tau) = \int_0^1 K(\tau) c\tilde{}(\tau) d\tau$$

One can often write approximately

$$C\tilde{}(K) = (1/n)\sum_{j=1}^{n} K((j-.5)/n) c\tilde{}((j-.5)/n)$$

The score function is usually chosen as a sequence of Orthonormal functions $\psi_1(.), \psi_2(.), \ldots$, especially the Legendre polynomials, which test against patterns in the change density $c\tilde{}(\tau)$.

The key to change analysis is to choose transformations of data (score the data) which are most powerful for detecting change. From the sample change processes, suitable linear

functionals (score the change) are formed. These linear functionals are called "double score components". One can define bivariate density functions $d(\tau, u)$, $0 < \tau < 1, 0 < u < 1$, of which double score functions are diagnostics. Choice of data score functions are motivated in sections 8 and 7 parametrically and non-parametrically, respectively.

Phase 4: *Density estimation.* By one of the many methods available in the theory of curve smoothing (kernel methods, splines, exponential methods, wavelets, etc.) form a smooth estimator $\hat{c(\tau)}$ of the change density.

An exposition of the theory of these phases would require a book and is beyond the scope of this paper. Our goal in this paper is to outline the phases and to explain how we choose transformations of the original data from which to form a change process.

7. NONPARAMETRIC STATISTICS MULTI-SAMPLE ANALYSIS

To test the equality of c samples non-parametric statistics starts by transforming each observation Y_{ji} to its "mid-rank" in the pooled sample. Let $F_{Y}\tilde{}$ and $p_{Y}\tilde{}$ denote the sample distribution and probability mass functions in the pooled sample. Define the mid-distribution function

$$F_{Y}\tilde{}^{mid}(y) = F_{Y}\tilde{}(y) - .5p_{Y}\tilde{}(y).$$

Let $J(u), 0 < u < 1$, be a score function. Transform Y_{ji} to

$$Z_{ji} = J(F_{Y}\tilde{}^{mid}(Y_{ji})).$$

Our definition of transformed data Z handles tied data and discrete data without extra effort. Traditional definitions assume all values in the pooled sample are distinct, and transform Y_{ji} to

$$Z_{ji} = J(nF_{Y}\tilde{}(Y_{ji})/(n+1)) = J(R_{ji}/(n+1)),$$

where R_{ji} is the rank in the pooled sample of Y_{ji}.

We calculate for the transformed data Z the correlation type statistic R_Z^2 from Z values in exactly the same way that R^2 was calculated from Y values. Asymptotically for $J(u) = u$, $S_Z = 1/12$, so that we could define

$$R_Z^2 = 12\sum_{j=1}^{c} p_j(Z\bar{}_{j.} - Z\bar{}_{..})^2$$

Note $Z\bar{}_{..} = .5$. The Kruskal-Wallis statistic equals $(n+1)R_Z^2$,

$$R_Z^2 = 12((\sum_{j=1}^{c}(n_j/n)Z\bar{}_j^2 - .25),$$

where $Z\bar{}_j$ is the rank average in the j-th group, traditionally computed

$$Z\bar{}_j = (1/n_j)\sum_{i=1}^{n_j} R_{ji}/(n+1)$$

Traditional non parametric computes only numerical test statistics such as R^2 corresponding to a score function $J(u) = u$. The change analysis approach to non-parametric analysis with score function $J(u) = u$ starts with a change density defined by

$$c_{\tilde{Z}}(\tau) = 12^{.5}(\tilde{Z}_{j.} - .5), \tau_{j-1} < \tau < \tau_j,$$

does graphical data analysis of its change process $\tilde{C_Z}(\tau)$, and computes double score components $[K, J]$.

8. FISHER-SCORE CHANGE PROCESSES

To detect change over time in a sequence one must have some prior opinion about the ways in which the probability distribution of the observations may be changing (such as in location, scale, skewness, etc). Sample change processes are formed for transformed data, where the transformation is called intuitively a *data score* function. The most powerful data transformations are essentially the sufficient statistics, or more precisely the Fisher score functions, when one has a parametric model $f(y;\theta)$ for a random sample $Y(t), t = 1, \ldots, n$, where $\theta = (\theta_1, \ldots, \theta_k)$.

The maximum likelihood estimator $\hat{\theta}$ is obtained by maximizing the average log-likelihood

$$L(\theta) = (1/n)\sum_{t=1}^{n} \log f(Y(t);\theta)$$

Define score functions

$$S_j(Y;\theta) = \partial/\partial\theta_j \log f(Y;\theta)$$

The maximum likelihood estimator is the solution of the *estimating equations* for $j = 1, \ldots, k$

$$(1/n)\sum_{t=1}^{n} S_j(Y(t);\hat{\theta}) = 0.$$

Our approach to change analysis asks if for every potential changepoint $\tau = m/n$ the parametric model with $\theta = \hat{\theta}$ fits the data $Y(t)$, $t = 1, \ldots, m$, up to the time m in the sense that approximately

$$(1/n)\sum_{t=1}^{m} S_j(Y(t);\hat{\theta}) = 0.$$

We define the *Fisher-score change* process to linearly interpolate its values at $\tau = m/n$, for $m = 1, \ldots, n$

$$\tilde{C}(\tau; S_j) = (1/n)\sum_{t=1}^{m} S_j^*(Y(t);\hat{\theta})$$

where

$$S_j^*(Y;\hat{\theta}). = S_j(Y;\hat{\theta})/E_{\hat{\theta}}^{.5}[S_j(Y;\hat{\theta})].$$

We form k Fisher-score change processes, for $j = 1, \ldots, k$.

We call this approach "random walk (or CUSUM) your normalized scores." We are developing the probability theory of the Fisher-score change processes.

These theoretical concepts can best be understood through examples. Consider a gamma distribution model

$$f(y;\nu,\theta) = (\theta^{\nu}\Gamma(\nu))^{-1} x^{\nu-1} \exp(-y/\theta)$$

where θ is a positive scale parameter, assumed unknown, and ν is a positive shape parameter, assmed known. One can show that the score function of the parameter θ is

$$S(Y;\theta) = (1/\theta)((Y/\theta) - \nu);$$

the maximum likelihood estimator is

$$\theta\hat{} = Y^{-}/\nu;$$

the normalized score function evaluated at the maximum likelihood estimator of the parameter may be shown to be

$$S^*(Y(t);\hat{\theta}) = \nu^{.5}((Y(t)/Y^{-}) - 1).$$

To test the observations $Y(.)$ for change, one forms the maximum likelihood score change process $C^{\sim}(\tau; S^*)$, $0 < \tau < 1$, and tests if this dynamic statistic is significantly different from a sample path of a Brownian Bridge stochastic process. A linear functional of the change process corresponding to the score function

$$K(\tau) = 12^{.5}(\tau - .5)$$

is

$$C^{\sim}(K, S^*) = (1/n)\sum_{t=1}^{n}(12\nu)^{.5}((Y(t)/Y^{-}) - 1)((t - .5)/n)$$

$$= (12\nu)^{.5}(1/n)\sum_{t=1}^{n} Y(t)((t - .5)/n)/Y^{-}$$

Under the null hypothesis of no change the asymptotic distribution of $n^{.5}C^{\sim}(K, S^*)$ is Normal(0,1).

An example of an application of this statistic is in Hsu (1979) where it is presented as a test designed for a small change in the scale parameter θ of an independent Gamma distributed sequence, derived by Kander and Zacks (1966) by a Bayesian analysis assuming the changepoint τ is uniformly distributed in time. This test statistic is derived in our approach as analogous to a component in standard goodness of fit analysis.

REFERENCES

Akaike, H. (1985). Prediction and entropy, *A Celebration of Statistics*, A. C. Atkinson and S. E. Feinberg, eds. Springer Verlag: New York, 1-24.

Alexander, William Pyle. (1989). *Boundary kernel estimation of the two sample comparison density function.* Ph.D. Thesis, Department of Statistics, Texas A&M University.

Aly, Emad-Eldin A. A., Miklós Csörgő, and Lajos Horváth. (1987). P-P plots, rank processes, and Chernoff-Savage theorems, in *Applied Statistics* edited by Madan L. Puri, Jose Perez Vilaplana and Wolfgang Wertz, New York: John Wiley & Sons, Inc.

Basseville, M. and A. Benveniste, eds. (1986). *Detection of Abrupt Changes in Signals and Dynamical Systems*, New York: Springer Verlag.

Blum, J.R., J. Kiefer and M. Rosenblatt. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist* 32, 485-498.

Carlstein, E. (1988). Nonparametric change-point estimation, *The Annals of Statistics.* 16, 188-197.

Csörgő, Miklós and Lajos Horváth. (1987). Nonparametric tests for the changepoint problem. *Journal of Statistical Planning and Inference*, 17, 1-9.

Csörgő, Miklós and Lajos Horváth. (1988). Nonparametric methods for changepoint problems. *Handbook of Statistics*, Vol. 7, P. R. Krishnaiah and C. R. Rao, eds. 403-425.

Eubank, R. L., V. N. LaRiccia, and R. B. Rosenstein. (1987). Some new and classical tests derived as components of Pearson's phi-squared distance measure *J. Amer. Statist. Assoc.* 82, 816-825.

Horváth, Lajos and Edit Gombay. (1990). Asymptotic distributions of maximum likelihood tests for change in the mean. *Biometrika*, 77, 2, pp. 411-4.

Hsu, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis, *Journal of the American Statistical Association*, 74, 31-40.

Kander, Z., and Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points, *Annals of Mathematical Statistics*, 37, 1196-1210.

Kolmogorov, A. N., Yu. V. Prokhorov, and A. N. Shiryaev. (1990). Probabilistic-statistical methods of detecting spontaneously occurring effects. *Proceedings of the Steklov Institute of Mathematics*, Issue 1, 1-21.

Krishnaiah, P. R. and B. Q. Miao. (1988). Review about estimation of change points. *Handbook of Statistics*, Vol. 7, P. R. Krishnaiah and C. R. Rao, eds. 375-402.

Müller, Hans-Georg and Jane-Ling Wang. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika*, 77, 2, pp. 305-14.

Parzen, Emanuel. (1979). Nonparametric statistical data modeling, *Journal of the American Statistical Association*, (with discussion), 74. 105-131.

Parzen, Emanuel. (1983). FUN.STAT quantile approach to two sample statistical data analysis. Technical Report. Invited paper, Canadian Statistical Society 1983 meeting in Vancouver.

Parzen, Emanuel. (1989). Multi-sample functional statistical data analysis, in *Statistical Data Analysis and Inference Conference in Honor of C. R. Rao*, ed. Y. Dodge, Amsterdam: Elsevier, 71-84.

Parzen, Emanuel. (1991). Goodness of fit tests and entropy, *Journal of Combinatorics, Information, and System Science*, 16, 129-136.

Parzen, Emanuel. (1991). Unification of statistical methods for continuous and discrete data, *Proceedings Computer Science-Statistics INTERFACE '90*, (ed. C. Page and R. LePage), Springer Verlag: New York, 235-242.

Parzen, Emanuel. (1992). Comparison change analysis. *Nonparametric Statistics and Related Topics* (ed. A. K. Saleh), Elsvier: Amsterdam, 3-15.

Parzen, Emanuel. (1993). From comparison density to two sample data analysis, *The Frontiers of Statistical Modeling: An Informational Approach*, ed. H. Bozdogan, Kluwers, Amsterdam.

Ruymgaart, F. H. (1974). Asymptotic normality of non-parametric tests for independence. *Ann. Statistics*, 2, 892-910.

Sen, P. K. (1981). *Sequential Nonparametrics.* New York: Wiley.

Tilksnys, Laimutis, ed. (1986). *Detection of changes in random processes.* Optimization Software, Inc., New York.

Weiss, L. (1964). On the asymptotic joint normality of quantiles from a multivariate distribution. *Jour. Research Nat. Bur. Standards B*, 68, 65-66.